# Assessing the cartographic performance of real-time quadtree-based generalisation of point data

Bereuter, Pia ; Weibel, Robert

# Assessing the Cartographic Performance of Real-Time Quadtree-based Generalisation of Point Data

P. Bereuter[1], R. Weibel[1]

[1]Department of Geography, University of Zurich, Winterthurerstrasse 190, 8057 Zürich
Email: {pia.bereuter | robert.weibel}@geo.uzh.ch

## 1. Introduction

With the advent of mobile and web mapping applications in recent years, generalisation for web and mobile mapping has gained importance. In this context real-time generalization techniques are a prerequisite for generalisation (Weibel and Burghardt 2008) in order to facilitate dynamic interaction and adaptation to the user. Typical applications, such as mashups or location-based services (LBS), usually encompass a thematic *foreground layer* predominantly in the form of points of interest (POIs) or large point collections (e.g. animal observations or twitter counts), against a spatial reference formed by *background map* such as a topographic map.

So far, the generalisation of point features has received only limited attention in map generalisation research. However, with the significant role of POI data in mobile apps and LBS and the predominance of point geometries in web and mobile applications, the development of generalisation algorithms for point features, capable of operating in real-time, can no longer be neglected by generalisation research.

This paper presents the continuation of previous work on investigating and developing real-time generalisation solutions for point data in web and mobile mapping applications. These real-time algorithms, described in Bereuter and Weibel (2012, 2013), use a quadtree as a spatial index data structure, since quadtrees both inherently reflect the spatial distribution of point objects and also allow to achieve efficient algorithms. We provide examples of the usage of the algorithms by Bereuter and Weibel (2012, 2013), and conduct a quantitative cartographic analysis using lichens data for Switzerland. The analysed algorithms perform in real-time on a prototype map client, with an integrated cartographic analysis module (Bereuter and Weibel 2011). The map client also supports dynamic map generalisation, and implements *content zooming* (Bereuter et al. 2012). The implemented algorithms are generic and not restricted solely to map clients. However, architectural considerations of client vs. server-side computing are not the issue of this paper.

## 2. State of the Art

Detailed reviews of real-time generalisation algorithms can be found, for instance, in (van Oosterom 2005, Weibel and Burghardt 2008, Bereuter and Weibel 2010, van Oosterom and Meijers 2011). In summary of these reviews, existing real-time methods either rely on pre-computation and storage in hierarchical data structures, or on generalisation algorithms that are sufficiently efficient so they can achieve real-time performance. While the first approach as a consequence of pre-computation lacks flexibility, the latter commonly sacrifices cartographic quality to reduce computational complexity.

Approaches of the first type that uses pre-computation are often based on dynamic or static tree data structures, to facilitate real-time generalisation (Cecconi and Galanda 2002, de Berg et al. 2004, Sester and Brenner 2004, 2008, Follin et al. 2005, van

Oosterom 2005, Haunert et al. 2009). Initial ideas to use quadtrees and hierarchical drainage basins to support generalisation have been proposed by Burghardt et al. (2004) and Edwardes et al. (2005). Dutton (Dutton 1999) developed a space-efficient, quadtree-like encoding scheme for positions on the globe called the quaternary triangular mesh (QTM), and used it for line filtering in map generalisation.

Point data generalisation algorithms of the second type include simple heuristics for selection, simplification, aggregation, typification and displacement of point sets (Openshaw 1995, Glover and Mackaness 1999, Edwardes *et al.* 2005, Lehto and Sarjakoski 2005). Of these map generalisation operators, algorithms for typification and displacement (Mackaness and Purves 2001, Regnauld 2001, Cecconi and Galanda 2002, Burghardt et al. 2004, Burghardt and Cecconi 2007) tend – since they typically rely on iterative optimisation procedures – to be more costly and hence are often avoided in a real-time environment.

The investigated algorithms in this paper, relying on a quadtree as an auxiliary data structure (Bereuter and Weibel 2013), bridge the dichotomy between approaches based on pre-computation and fast generalisation algorithms, that is between flexibility and performance.

## 3. Methods and Data

To resolve spatial conflicts in map generalisation different generalisation operators are applied. McMaster and Shea (McMaster and Shea 1992) presented a topology for generalisation operators in line generalisation. An approach to specifically classify point generalisation algorithms has been proposed by Bereuter and Weibel (2010).

The algorithms described in Bereuter and Weibel (2013) are all based on the use of the quadtree data structure and provide realizations of the major generalization operators that can be applied to point data. Several algorithms are available for each of the generalization operators summarized in (Table 1). The basic idea of the quadtree-based generalisation approach, to apply the generalisation operators on the quadtree nodes (short: quadnodes) according to the target level of detail (LOD), which is mapped to the depth of the quadtree. LOD translates to the width of the quadnode side, measured in screen (pixel) coordinates and denotes the smallest required distance to resolve spatial conflicts, given cartographic constraints (e.g. point symbol size or smallest separation distance between point symbols).

Table 1. Quadtree based generalization operators (cf. Bereuter and Weibel, 2013).

| Operator | Description |
|---|---|
| Selection | Based solely on feature attributes; applying various selection functions, such as rank, frequency or feature category distribution. |
| Simplification | Returns one point feature per quadnode, governed by geometric criteria such as centrality, or weighted centrality within the quad. |
| Aggregation | Reduces the number of points by grouping together semantically similar or spatially close points, replacing the original points by a new placeholder feature, such as midpoint, clustering, or collocation occurrence. |
| Displacement | Locally reconfigures point symbols to resolve spatial conflicts by moving points apart from each other. Uses the quadtree spatial data structure for neighbour search. |

As the focus of this paper lies on the cartographic analysis of the generalisation of these algorithms and due to limited space we refer to Bereuter and Weibel (2013) for further details on the algorithms and on their performance. A prototype map client was implemented that provides the real-time generalisation algorithms based on Java and Processing (www.processing.org). In addition to the point generalisation module and the support of selected map services, the development environment contains a module for the multiscale analysis of the resulting generalised point data sets. Based on this client Bereuter et al. (Bereuter et al. 2012) have shown examples of how the different operators can feed into a workflow and introduced the concept of *content zooming*, which offers the user a possibility to change the amount and granularity of foreground information presented, without changing the geometric map scale.

The dataset (Figure 1) used for the cartographic analysis of the generalisation algorithms is a point collection of lichens observations originating from SwissLichens (Stofer et al. 2012), a database recording past and present population distribution of more than 500 different lichen species at close to 87,000 locations within Switzerland. This dataset is particularly interesting as a large thematic dataset to analyse real-time generalisation output in terms of quality, as it comes with a comparably large spatial extent and features a large set of categories and attributes on each lichen location, as well as information about data precision, time and several indicator values.

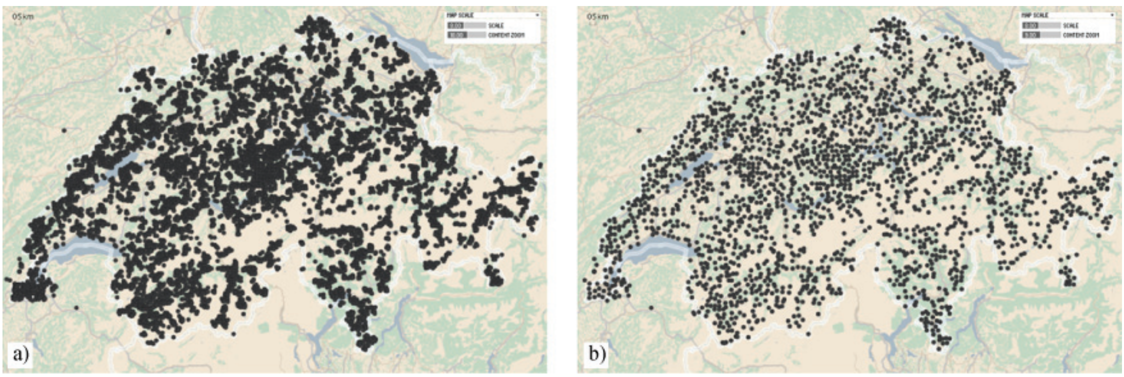


Figure 1. Lichens observations in Switzerland. (a) Raw data from SwissLichens. (b) Overview at zoom level 8, using centrality-based simplification.

## 4. Analysis and Discussion

In this section we show a quantitative analysis of the results that can be obtained using the algorithms described in Bereuter and Weibel (2013). Due to space limitations, the analysis is restricted to selected algorithms, and to the main cartographic requirements guiding map generalisation:

- Data reduction: What is the reduction rate in relation to the zoom level?
- Conflict reduction: How many overlaps between points symbols are resolved?
- Data enhancement: How are important point attributes retained?
- Displacement measures: How is displacement achieved?
- Maintenance of spatial patterns: Is the overall point distribution maintained?

### 4.1 Data Reduction

Figure 2 shows the data reduction curves for the quadtree-based algorithms, as well as for the Radical Law (Töpfer and Pillewizer 1966). Not surprisingly, point reduction algorithms (selection, simplification, aggregation) retain less points than displacement, for those zoom levels where most conflicts arise. It also becomes clearly visible how the quadtree-based algorithms, at any zoom level, retain more points than the Radical Law would suggest. This dramatically different behavior may be surprising, but is due to different underlying principles of point reduction used. The Radical Law was originally empirically developed using topographic map series; hence it does indeed reflect a good first estimate of the objects to be retained. However, when used as measure to guide the generalisation process, the spatial arrangement of point features must be controlled by the generalisation algorithm applied, since the Radical Law merely uses the ratio of the square roots of scale denominators and thus has no concept of space *per se*. The quadtree-based generalisation algorithms, on the other hand, inherently take the proximity and density of point symbols into account, owing to the quadtree structure. These algorithms will only remove or aggregate points if spatial conflicts occur. Thus, between levels 20 and 15, almost no point reduction takes place. Once the average distance between data points reaches the size of the quad cell corresponding to the given zoom level, however, the point reduction rate rapidly increases, due to the working principle of the quadtree-based algorithms. At this point the slope of the data reduction curves becomes even steeper than the one of the Radical Law.

In Figure 2 the data reduction of the generalisation operators applied to the SwissLichens dataset is also set into relation with three different artificially generated datasets with the same number of points, and applied to the same spatial extent (given by the axis-aligned bounding box). The more regular the spatial distribution, the steeper the S-shape of the curves, due to the uniformity of the arrangement. A similar but less distinct curve progression is shown by the random point pattern, as it does not contain distinct clusters that are reduced by the generalisation operators at higher zoom levels. Finally, two samples of clustered point patterns are also plotted in Figure 2. The generated patterns differ in the number of clusters, their size and the attributed standard deviation around the cluster centre. Dependent on their parameterisation the S-curve of the data reduction shifts its inflection point to higher or lower zoom levels.

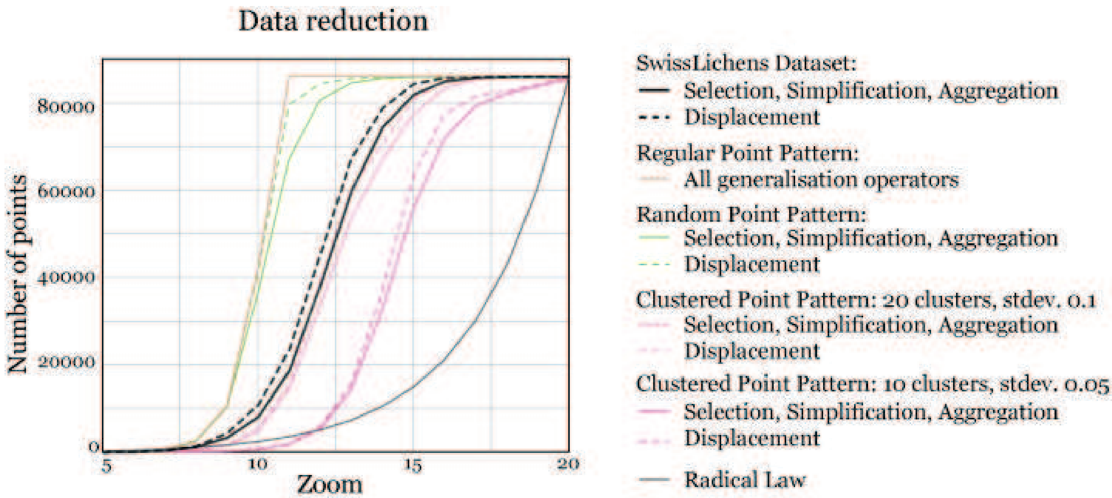


Figure 2. Data reduction for different datasets with the same spatial extent and number of points; SwissLichens, and artificial regular, random and clustered point patterns.

### 4.2 Conflict Reduction

Those zoom levels with a high data reduction in Figure 2 hint to a high number of cartographic conflicts, that is, overlapping symbols, at those zoom levels. Figure 3 shows the evolution and reduction of cartographic conflicts over the different zoom levels for the selection algorithm, with the different variations of collision checks applied. The point reduction algorithms (selection, simplification, aggregation) only retain a single point per quadnode for the target zoom level (Bereuter and Weibel 2012, 2013).

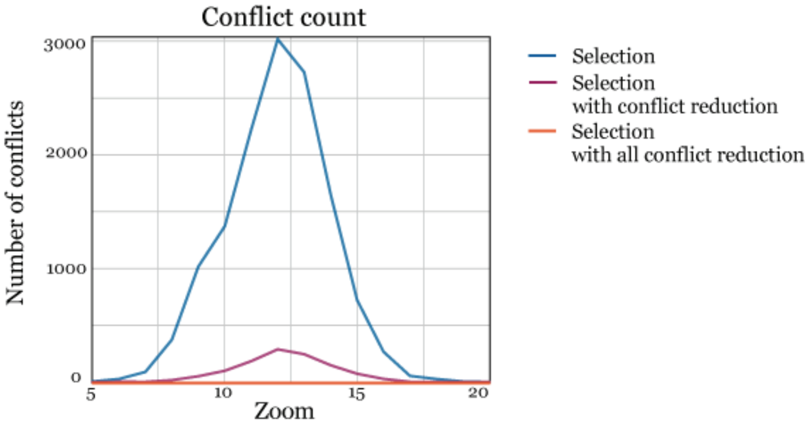


Figure 3. Conflict counts for selection without conflict checks, and selection with different forms of conflict checks applied. Number of points in SwissLichens: 86,845.

However, cartographic conflicts are not entirely removed by solely retaining one point per quadnode, as this does not consider potential overlaps of points contained in neighbouring quadnodes (Figure 3, blue curve). That is, two points might be lying across the border between two neighbouring quadnodes, separated by a distance less than the symbol size. This can be alleviated by checking for collisions in the four – respectively eight possible quadnode neighbours (Figure 3, red line), or by moving all points towards the centre of the quadnode, not allowing for any overlap (Figure 3, orange line).

Figure 4 shows results applying the different variants of conflict reduction checks in the case of value-based selection of lichens data for the SW part of Switzerland – the lower part of the Swiss Rhone valley. While pure selection still shows a considerable number of cartographic conflicts (red dots in Figure 4a), the use of conflict constraints in the algorithm reduces them significantly (Figure 4b and 4c). Moving the points to towards the centre of the quadnode is faster performance-wise compared to checking for conflicts in the neighbouring quadnode, but introduces regularity to the generalised dataset (not shown in Figure 4).

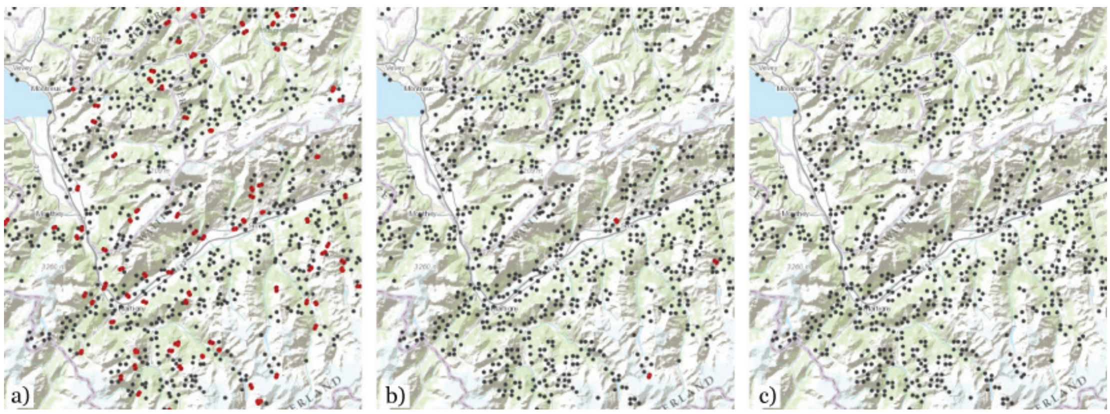


Figure 4. Selection operator at zoom level 10: a) without conflict reduction, b) with horizontal and vertical neighbour conflict checks, c) including corner checks.

### 4.3 Data Enhancement

Quadtree-based selection allows to retain peculiarities of the point attributes across zoom levels. In Figure 5 a variant of value-based selection – selection based on local maxima of the attribute – illustrates how a particular attribute is retained through a series of scales. In this case, selecting the points with maximum value for the red list status favours most endangered species (red dots), rather than maintaining the overall distribution of categories across the various zoom levels. Retaining the average of lichens per quadnode would basically only show green spots due to the skewed distribution of common lichens as opposed to endangered lichen species. Selection based on the local attribute value distribution maintains to a certain extent the underlying spatial distribution of the point pattern, and highlights clusters where in this case most endangered lichens are located. Each location with occurrence of lichens therefore shows the most endangered lichen per quadnode. That is, if at that location only less endangered lichens exist, the most endangered representative (green or yellow dot) of them is shown. Local selection therefore effectively shows the local context among endangered lichens as opposed to only representing most endangered lichens (Figure 5d). However, global hotspots are less pronounced by this type of selection, as can be observed by comparing Figure 5c and 5d.

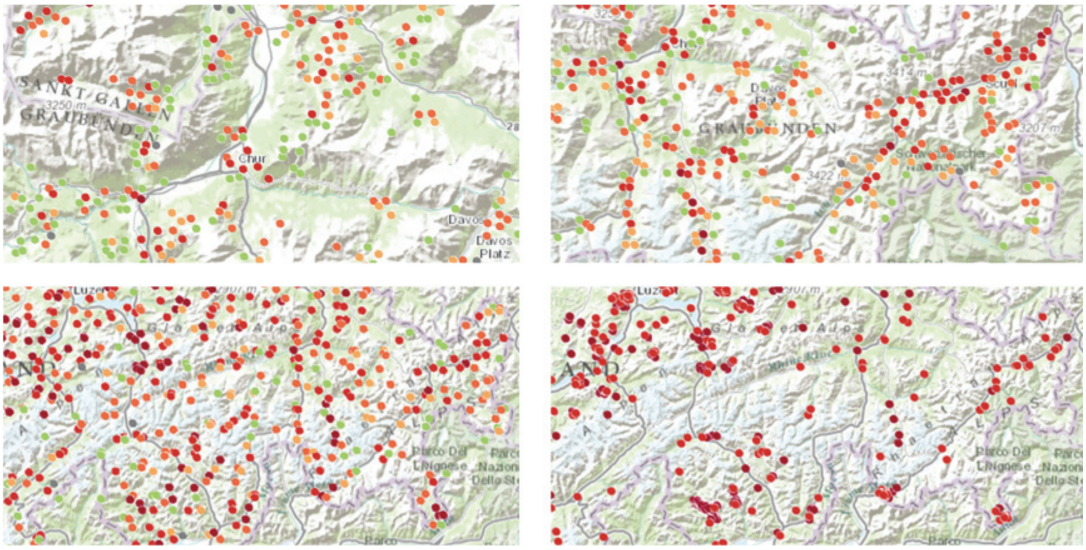


Figure 5. Local value-based selection, retaining most endangered lichens, from zoom level 10 (a) to zoom level 8 (c). d) Global value-based selection. Colours range from least endangered in green to most endangered lichen species in red.

### 4.4 Displacement Measures

Displacement algorithms help to further resolve spatial conflicts and thus allow retaining more point symbols than solely with point reduction operators. The displacement algorithm tries to accommodate as many points as possible keeping at most one point per quadnode, and displacing points if the neighbouring quadnodes provide sufficient holding capacity for displacement (Bereuter and Weibel 2012, 2013). Remaining overlaps can be removed by further solving boundary constraints as illustrated in Figure 4. The comparison of the results of mere point reduction (by centrality-based simplification) vs. the displacement algorithm (Figure 6 a,b) illustrates that displacement retains more points for the displayed zoom level. On the other hand, it shows that displacement has the effect of homogenizing dense clusters – especially if only horizontal and vertical quad neighbours are considered – and thus affecting the overall distribution pattern. The characteristics of the displacement operator applied can be highlighted by plotting cumulative displacement vectors (in number of pixels) for each angle of displacement. Figure 6c shows nicely that the algorithm – considering only horizontal and vertical neighbours – did not displace points to diagonal neighbours with the majority of displacement angles arranged in the four cardinal directions.

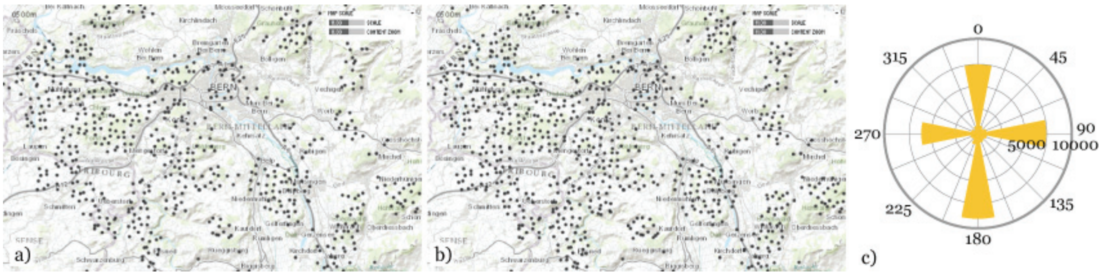


Figure 6. a) Centrality-based simplification (647 points). b) Displacement applied after centrality-based simplification (788 points). c) Cumulative displacement vectors for zoom level 11 for the region around the city of Bern in Switzerland.

### 4.5 Maintenance of Spatial Patterns

How well a generalization algorithm maintains the underlying spatial point distribution pattern can be investigated by visually comparing the kernel density estimation (KDE) of a point pattern for different zoom levels, or analysed quantitatively, by calculating the difference between two kernel density estimations. The kernel density map in Figure 7a for zoom level 9, and the KDE difference map in Figure 7b between zoom levels 9 and 10, show where the hotspots of the density distribution of the points are located and to what degree the point pattern changed, respectively, giving a locally differentiated picture. It shows that the applied algorithm reduces the point density most at local peaks, where the highest densities are located and that overall, it maintains the underlying spatial pattern. It also shows that the algorithm hardly changes the point density in regions of low density, therefore the overall maxima of the point distribution is less evident. This effect can be reduced by applying a different parameterisation of the generalisation operator.

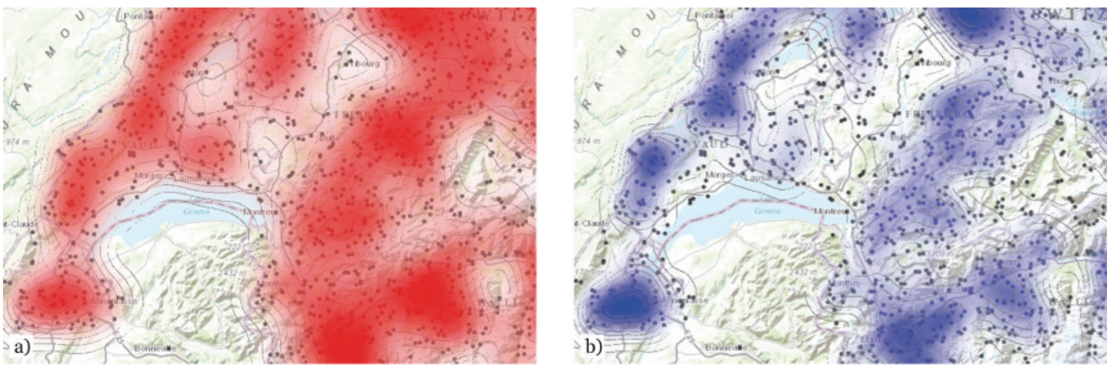


Figure 7. a) Kernel density estimation for selection at zoom level 9, b) Kernel density difference between zoom levels 9 and 10.

## 5. Conclusion and Outlook

In previous research we introduced a set of object-directed, real-time generalisation algorithms for point data, providing solutions for the major generalisation operators – selection, simplification, aggregation and displacement (Bereuter and Weibel 2012, 2013). These algorithms use a quadtree as an auxiliary index structure designed for speed to achieve on-the-fly generalisation, and produce appropriate cartographic results. This paper presented the quantitative cartographic analysis of these algorithms applied in a set of experiments to the SwissLichens database (Stofer et al. 2012) featuring close to 87,000 lichen observations reported in Switzerland to date.

The cartographic analysis covers selected aspects of the generalisation results, some strengths and weaknesses. The analysis of data reduction applied to different datasets shows that the data reduction curve does not follow the Radical Law (Töpfer and Pillewizer 1966), but rather follows an S-shape, due to the underlying spatial distribution of the data, that is not directly reflected in the formula of the Radical Law. The data reduction curves further indicate in their steepest part at which zoom level most cartographic conflicts occur, which is confirmed by analysing the occurrence and reduction of cartographic conflicts per zoom level. Comparing local and global selection criteria indicates that through local selection and data enhancement the context of the spatial distribution is maintained. A comparison between simplification and displacement operators demonstrates the additional number of points that can be accommodated by the displacement operator. And finally, the analysis of the spatial point distribution shows that the spatial arrangement is mainly changed at locations with high point density. Low density regions may be overrepresented, which can be tackled by a different parameterisation of the algorithm.

As future work, the thorough analysis based on different use cases is currently under way, the detailed results of which will be reported in the thesis. A further step involves integrating and testing of space-directed generalisation (Bereuter and Weibel 2010) in the quadtree-based environment, using a density-equalizing deformation algorithm. Beyond the scope of this project, future research may explore potential improvements of the current algorithms, such as the inclusion of spatial constraints imposed by the background map, as well as more elaborate displacement strategies.

## Acknowledgements

The research reported in this paper represents part of the PhD project of the first author. Funding by the Swiss National Science Foundation through the project Generalisation for Portrayal in Web and Wireless Mapping (GenW2+) (grant no. 200020-138109) is gratefully acknowledged. Particular thanks go to Silvia Stofer of the National Data and Information Center for Swiss lichens and all the contributors for collecting and providing the data of the SwissLichens dataset.
Image sources: Base map © 2013 ESRI ArcGIS Services World Topo Map

## References

Bereuter P and Weibel R, 2010. Generalisation of point data for mobile devices: A problem-oriented approach. *In*: S. Mustiere and W.A. Mackaness, eds. *13th Workshop on Progress in Generalisation and Multiple Representation*. Zurich: ICA Commission on Generalisation and Multiple Representation, 1–8.

Bereuter P and Weibel R, 2011. A Diagnostic Toolbox for Assessing Point Data Generalisation Algorithms. *In*: *Proceedings of the 25th International Cartographic Conference*. Paris, France, 1–10.

Bereuter P and Weibel R, 2012. Algorithms for on-the-fly generalization of point data using quadtrees. *In*: *Proceedings of the 17th AutoCarto conference*. Columbus, OH, 16.

Bereuter P and Weibel R, 2013. Real-time generalization of point data in mobile and web mapping using quadtrees. *Cartography and Geographic Information Science*, 40 (4), 1–11.

Bereuter P, Weibel R and Burghardt D, 2012. Content zooming and exploration for mobile maps. *In*: J. Gensel, D.J. And, and D. Vandenbroucke, eds. *Proceedings of the AGILE 2012 International Conference onf Geographic Information Science*. Avignon, 74–80.

De Berg M, Bose P, Cheong O and Morin P, 2004. On simplifying dot maps. *Computational Geometry*, 27 (1), 43–62.

Burghardt D and Cecconi A, 2007. Mesh simplification for building typification. *International Journal of Geographical Information Science*, 21 (3), 283–298.

Burghardt D, Purves RS and Edwardes AJ, 2004. Techniques for on the-fly generalisation of thematic point data using hierarchical data structures. *In*: *Proceedings of the GIS Research UK 12th Annual Conference*. Norwich, UK.

Cecconi A and Galanda M, 2002. Adaptive zooming in Web cartography. *In*: *Computer Graphics Forum*. Amsterdam, Holland, 787–799.

Dutton G, 1999. Scale, Sinuosity, and Point Selection in Digital Line Generalization. *Cartography and Geographic Information Science*, 26 (1), 33–54.

Edwardes AJ, Burghardt D and Weibel R, 2005. Portrayal and Generalisation of Point Maps for Mobile Information Services. *In*: T. Reichenbacher, L. Meng, and A. Zipf, eds. *Map-based mobile services: theories, methods and implementations*. Berlin Heidelberg: Springer, 11–28.

Follin J, Bouju a, Bertrand F and Boursier P, 2005. Multi-resolution extension for transmission of geodata in a mobile context. *Computers & Geosciences*, 31 (2), 179–188.

Glover E and Mackaness WA, 1999. Dynamic generalisation from single detailed database to support web based interaction. *In*: *Proceedings 19th Int. Cartographic Conference*. Ottawa, Canada, 1175–1184.

Haunert J-H, Dilo A and van Oosterom P, 2009. Constrained set-up of the tGAP structure for progressive vector data transfer. *Computers & Geosciences*, 35 (11), 2191–2203.

Lehto L and Sarjakoski LT, 2005. Real-time generalization of XML encoded spatial data for the Web and mobile devices. *International Journal of Geographical Information Science*, 19 (8-9), 957–973.

Mackaness WA and Purves RS, 2001. Automated Displacement for Large Numbers of Discrete Map Objects. *Algorithmica*, 30 (2), 302–311.

McMaster RB and Shea KS, 1992. *Generalization in digital cartography*. Washington: Association of American Geographers Washington, DC.

Van Oosterom P, 2005. Variable-scale Topological Data Structures Suitable for Progressive Data Transfer: The GAP-face Tree and GAP-edge Forest. *Cartography and Geographic Information Science*, 32 (4), 331–346.

Van Oosterom P and Meijers M, 2011. Towards a true vario-scale structure supporting smooth-zoom. *In*: *14th ICA/ISPRSWorkshop on Generalisation and Multiple Representation & the ISPRS Commission II/2 WG on Multiscale Representation of Spatial Data*. Paris, 19.

Openshaw S, 1995. Developing Automated and Smart Spatial Pattern Exploration Tools for Geographical Information Systems Applications. *The Statistician*, 44 (1), 3–16.
Regnauld N, 2001. Contextual Building Typification in Automated Map Generalization. *Algorithmica*, 30 (2), 312–333.
Sester M and Brenner C, 2004. Continuous generalization for visualization on small mobile devices. *In*: O. Altan, ed. *Proc. Conference on Spatial Data Handling*. Istanbul: Springer, 469–480.
Sester M and Brenner C, 2008. A vocabulary for a multiscale process description for fast transmission and continuous visualization of spatial data. *Computers & Geosciences*, 35 (11), 2177–2184.
Stofer S, Scheidegger C, Clerc P, Dietrich M, Frei M, Groner U, Jakob P, Keller C, Roth I, Vust M and Zimmermann E, 2012. Nationales Daten- und Informationszentrum der Schweizer Flechten - SwissLichens. Birmensdorf, Schweiz: Datenbankauszug vom 25. September 2012 (SST_201209).
Töpfer F and Pillewizer W, 1966. The principles of selection. *Cartographic Journal*, 3 (1), 10–16.
Weibel R and Burghardt D, 2008. Generalization, On-the-Fly. *In*: S. Shekhar and H. Xiong, eds. *Encyclopedia of GIS*. New York: Springer, 339–344.
Yan H and Weibel R, 2008. An algorithm for point cluster generalization based on the Voronoi diagram. *Computers & Geosciences*, 34 (8), 939–954.